4. DATA

In California, there is no data-reporting requirement for self-insured employers that would provide the data necessary to conduct this study. This was the reason that self-insured claims were excluded from Peterson *et al* (1998). For the current study, RAND conducted an unprecedented data collection effort from self-insured employers in California, with assistance from the California Self-Insured Plans (their regulator), the California Self-Insurers Association (their lobbying group), and the Commission for Health and Safety and Workers' Compensation. The data collected from employers was then linked to wage data from the State of California to create a unique database of proprietary employer data linked to state administrative data. In this section, the self-insured data collection effort and administrative data link are described. More information is available on the self-insured data in Appendix A. This section also briefly describes the data on claims at insured firms in California, which are described in greater detail in Peterson *et al* (1998) and Reville (1999).

SELF-INSURED DATA

In response to the request of the Commission for Health and Safety and Workers' Compensation that RAND estimate earnings losses at self-insured employers in California, RAND contacted a sample of 150 private (out of 466) and 150 public (out of 432) self-insured firms and requested data on all indemnity claims from 1991-1996. The sample was based on the number of claims at the employer so that the resulting sample of claims was a representative sample of claims from self-insured employers (rather than a sample of self-insured firms), and the sample was stratified by employer size to increase the probability of selection for small self-insured employers.[1] Specifically, we requested data on benefit amounts paid and incurred, injury dates, and individual identifiers to facilitate linking to earnings data maintained by the State of California Employment Development Department (EDD). Concerned that requesting too many data elements would lead to a lower response rate, we also suggested additional data that could be provided optionally, including disability ratings, settlement method, and litigation indicators. The initial letter from RAND to the employers describing the data request was sent in May 1998. Accompanying this letter were letters encouraging participation from the California Self-Insured

[1] In addition, the sample was stratified to oversample employers that had not changed third-party administrators (TPA) since we were concerned that older data would not be available from TPAs who had not continuously serviced the employer.

Plans, the California Self-Insurers Association, and the Commission for Health and Safety and Workers' Compensation. A follow-up letter was sent in July 1998.

We received our data from the self-insured from June through August of 1998, with most of the data received during June and July. By June 15, we had received data from 74 companies. By July 1, we had received data from 107 employers. By July 15, we had received data from 143 employers. In several cases, problems were immediately identified at RAND when the data arrived, and new files were sent by the employers or the third-party administrators (TPA). By the end of August, we had received data from 167 employers, including 79 private employers. During this period, a RAND staff member fielded daily calls from employers and TPAs with questions, logged the arrival of incoming data, and organized the files on a secure computer so that the programmer could begin to process them.

Of the 79 data files from private, self-insured employers, 68 were included in the final sample. The remaining eleven were eliminated for various reasons including inability to identify PPD claims, inability to construct total indemnity, and lack of identifiers with which to link to EDD wage data. The 68 firms with data represent 15 percent of self-insured, private employers, and 30 percent of indemnity claims at self-insured employers.

Table 2 reports the final response rate broken down by industry, employer size quartile, and pre-injury earnings quintile. The top panel shows that the sample includes representation from every major industry category, though the likelihood of response differs considerably across industries. In particular, public utilities (communication, power, and water) were considerably more likely to respond, and transportation (primarily trucking) was less likely to respond. The middle panel shows a nonlinear relationship between firm size and the probability of response, though the largest firms were most likely to provide data, which led to some over-sampling by firm size. However, since smaller firms were more likely to be contacted, the net effect on firm size in the sample of self-insured claims, while still over-representing larger firms, is not large (see Table 1, comparing weighted and nonweighted samples). The bottom panel shows the response rate by quartile of per-employee payroll. Without conditioning on other variables, there is not a clear relationship between per-employee earnings and the probability of response. In the analysis below, estimates weighted to account for nonresponse and sampling are reported for the main results. See the appendix for further discussion of nonresponse and the construction of the weights.

INSURED DATA

For our analysis of PPD claimants at insured employers, we used claims data from the Workers' Compensation Insurance Ratings Bureau (WCIRB). The data are from the Uniform

Statistical Reporting Plan (USR) database from the WCIRB, a private entity responsible for proposing and publishing workers' compensation insurance premiums and class rates. All claims

Table 2

Response Rate by Industry and Firm Size

Industry	Number Sampled	Response Rate
Agriculture, Forestry and Fishing (SIC-0)	5	0.400
Mining and Construction (SIC-1)	5	0.200
Manufacturing (SIC-2)	28	0.357
Manufacturing (SIC-3)	33	0.424
Transportation (SIC-4)	8	0.125
Communication, Power, Water (SIC-4)	6	0.833
Retail, Wholesale Trade (SIC-5)	19	0.316
Financial, Hotels, Entertainment (SIC-6-7)	8	0.750
Health Care Services (SIC-8)	38	0.605
Total	150	0.453
Firm Size		
Less than 1040	30	0.367
1041-1832	30	0.467
1833-4098	30	0.533
4099-13127	30	0.333
13128 and greater	30	0.567
Total	150	0.453
Per-Employee Payroll		
Less than $17,180	30	0.533
$17,181-$23,274	30	0.400
$23,275-$31,896	30	0.467
$31,897-$42,096	30	0.400
$42,096 and greater	30	0.467
Total	150	0.453

for permanent partial disability in insured firms in California are reported to the WCIRB. We received data for claims that occurred on policies that opened from 1989 to 1994.[2] The data from the WCIRB provide detailed information about the characteristics of claims and injuries and benefits and expenses as they were incurred and paid, and some information about how claims were processed. These data are provided only for claims submitted against employers who are covered by workers' compensation insurance carriers. These are the same claims used for the analysis in Peterson *et al* (1997), though all of the claims information were updated in January 2000. As a result, the later accident years (1993-1995), which were not mature claims at the time of the first report, now reflect up to four years of development.

[2]Policies reopen every year, and therefore all policies with claims are included.

WAGE DATA

The wage data are from the Base Wage file maintained by EDD. Every quarter, employers covered by Unemployment Insurance (UI) in California are required to report the quarterly earnings of every employee to the EDD. These reports are stored in the Base Wage file. The industries covered by UI are virtually identical to the industries covered by workers' compensation,[3] therefore a worker injured at a firm for which he or she can make a workers' compensation claim should also have a record for that quarter in the Base Wage file.

There are several limitations of the EDD data. First, they do not report earnings in the uncovered sector, or, more importantly earnings in another state. The control methodology described above is partly intended to correct for this problem. Only if the injured worker is more likely than the control to receive earnings in the uncovered sector or out of state will this bias the result. Another limitation of the EDD data is the level of earnings reported, which is quarterly. For purposes of estimating total earnings loss, this is not a limitation, but with quarterly earnings data, it is impossible to distinguish between the hourly wage effects of a disability and a reduction in hours or weeks worked.

LINKING CLAIMS AND WAGE DATA AND SELECTING CONTROLS

The match rate of claims data to the wage data for the self-insured was very high. Out of 103,416 claims with individual identifiers for EDD matching provided by the employer, less than 2 percent (1,701) were not matched by Social Security number to the EDD data. A number of steps were then taken to arrive at the final analysis sample of 21,852 PPD claims from the second quarter of 1991 through the fourth quarter of 1995. Primarily, these steps involved dropping non-PPD (medical-only and temporary-only claims) claims, dropping claims after 1995 and before the second quarter of 1991, and selecting only first observed PPD claims. These steps are detailed in Appendix A.

To select controls for the injured workers, RAND provided EDD with a firm identifier that could be matched to the Base Wage file. EDD then identified all workers in the state who had worked at the employers at some point over the six years (1991-1996). EDD then created a database with quarterly earnings at every job in California from 1989-1998 for all workers at all 68 employers. After removing the injured workers from the database by using the individual identifiers provided by RAND, EDD stripped the identifiers for the uninjured workers from the data and provided the wage files to RAND. Using this data file, RAND was able to select up to

five controls for every injured worker. See the appendix Table A5 for a frequency table of the number of controls per injured worker at private, self-insured employers.

As discussed in Peterson et al (1997) and Reville (1999), the match rate of injured workers to the EDD wage data was also very high for the WCIRB data. Since receiving all earnings of every worker at every insured firm in California was not an option for the original study, a 20 percent random sample of claims was provided to EDD and EDD selected the controls. Approximately 65 percent of SSNs were matched to controls. The primary reason for a relatively low match rate of controls to injured workers is that small firms are less likely to have any other workers with wages in the allowed wage range,[4] which led to a sample of insured firms that overrepresents larger firms. See Peterson *et al* (1998) and Reville (1999) for more information on the insured sample.[5] See the appendix Table A4 for a frequency table of the number of controls per injured worker at the insured.

---

[3] In both systems, federal civilian and military employees, U.S. postal service workers, railroad employees, and the self-employed are excluded.

[4] EDD selected controls by choosing workers whose wages were within a fixed distance (approximately ten percent) of the wages of the injured worker. See the appendix of Reville (1999) for further information.

[5] A detailed description of the data is provided in the appendix of Reville (1999).